# DNA-binding mechanism of the *Escherichia coli* Ada *O*<sup>6</sup>-alkylguanine–DNA alkyltransferase

Philip E. Verdemato, James A. Brannigan<sup>1</sup>, Christian Damblon, Fabio Zuccotto<sup>2</sup>, Peter C. E. Moody<sup>\*</sup> and Lu-Yun Lian

Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK, <sup>1</sup>Department of Chemistry, University of York, Heslington, York YO10 5DD, UK and <sup>2</sup>Centre for Mechanisms in Human Toxicity, University of Leicester, Leicester LE1 9HN, UK

Received July 10, 2000; Revised and Accepted August 10, 2000

## ABSTRACT

The C-terminal domain of the Escherichia coli Ada protein (Ada-C) aids in the maintenance of genomic integrity by efficiently repairing pre-mutagenic O<sup>6</sup>alkylquanine lesions in DNA. Structural and thermodynamic studies were carried out to obtain a model of the DNA-binding process. Nuclear magnetic resonance (NMR) studies map the DNA-binding site to helix 5, and a loop region (residues 151-160) which form the recognition helix and the 'wing' of a helix-turn-wing motif, respectively. The NMR data also suggest the absence of a large conformational change in the protein upon binding to DNA. Hence, an O<sup>6</sup>-methylquanine (O<sup>6</sup>meG) lesion would be inaccessible to active site nucleophile Cys146 if the modified base remained stacked within the DNA duplex. The experimentally determined DNA-binding face of Ada-C was used in combination with homology modelling, based on the catabolite activator protein, and the accepted base-flipping mechanism, to construct a model of how Ada-C binds to DNA in a productive manner. To complement the structural studies, thermodynamic data were obtained which demonstrate that binding to unmethylated DNA was entropically driven, whilst the demethylation reaction provoked an exothermic heat change. Methylation of Cys146 leads to a loss of structural integrity of the DNA-binding subdomain.

## INTRODUCTION

Alkylation damage to DNA poses a threat to genomic integrity through the generation of pre-mutagenic base lesions. Such lesions include  $O^6$ -methylguanine ( $O^6$ meG) which can form a stable mispair with thymine. If left unchecked, such mispairing gives rise to a transition mutation. The  $O^6$ meG DNA-methyltransferase protein, Ada, represents the first line of defence against such lesions in *Escherichia coli*. This 354 residue protein is organised into two independently structured domains, each containing DNA repair activity.

The C-terminal domain transfers the alkyl group from the  $O^6$ meG to a cysteine thiol (Cys321) (1–4), in what is believed to be an S<sub>N</sub>2 mechanism (5–7). This transfer is stoichiometric, irreversible, and results in protein inactivation together with a regenerated guanine base.

The N-terminal domain of the Ada protein irreversibly transfers methyl groups from non-mutagenic methylphosphotriesters on the DNA backbone to its active site cysteine, Cys69. Methylation of this thiol triggers a conformational switch in the N-terminal domain of the protein, enabling it to bind specifically to the promoter regions of the *ada* gene, and a number of other genes involved in the response to alkylation damage (8-10). Once the Ada N-terminal domain is specifically bound to the promoter, the C-terminal domain causes transcriptional activation through contacts to the  $\sigma$ 70 subunit of RNA polymerase (11,12). The function of Ada as a transcription factor is thus dependent on the methylation of Cys69, and the presence of the C-terminal domain (regardless of methylation state) (8,13). This process of transcriptional enhancement is known as the 'adaptive response' and leads to further production of Ada protein capable of repairing alkylated DNA (14,15).

The mechanism by which the 19 kDa C-terminal domain of Ada (Ada-C) binds and repairs DNA has been the subject of much investigation. A helix-turn-helix motif was identified from the crystal structure of Ada-C (16). Mutation studies have implicated this helix-turn-helix region in binding DNA in the homologous human protein (17). The crystal structure of Ada-C also revealed the presence of a catalytic tetrad hydrogen bonding network involving an ordered water, reminiscent of the catalytic triad in serine proteases. The postulated role of this network was to deprotonate the cysteine thiol to generate the reactive thiolate anion for nucleophilic attack and methyl abstraction. Localisation of the active site thiol to the interior of the protein renders it inaccessible to an O<sup>6</sup>meG lesion stacked within a DNA duplex. In order to effect repair, it was suggested that the lesion flipped out of the double helix, and into the buried active site through an opening in the surface of the protein (18). A similar DNA-binding model has also been suggested for the human homologue (19). These DNA-binding

<sup>\*</sup>To whom correspondence should be addressed. Tel: +44 116 252 3366; Fax: +44 116 252 3473; Email: pcem1@le.ac.uk Present address:

Lu-Yun Lian, Department of Biomolecular Sciences, UMIST, Manchester M60 1QD, UK

models were largely based on mutation studies and not direct experimental evidence. We have used nuclear magnetic resonance (NMR) and other biophysical techniques to elucidate how Ada-C binds to a small DNA fragment. The data show that a helix–turn–wing motif (20) constitutes the DNA-binding site for both single-stranded (ss) and double-stranded (ds) DNA. When binding duplex DNA, Ada-C appears to predominantly contact a single DNA strand. Furthermore, the combined experimental and modelling data show that base flipping of the  $O^6$ meG lesion appears to be a functional necessity for repair to occur.

# MATERIALS AND METHODS

# Ada-C mutagenesis

Site-directed mutagenesis was performed using the Sculptor kit (Amersham) which is based on the use of oligonucleotides to direct mutation after priming on circular ssDNA, and selection against the parental strand using a modified Eckstein method (21). As an aid to mutant screening, restriction sites were incorporated that do not affect the amino acids encoded. Final constructs were sequenced to ensure no other mutation had occurred.

## Production of isotopically-labelled and unlabelled Ada-C

Clones of wild-type and C146S mutant Ada-C were expressed in B834(DE3) *E.coli* cells using a pET plasmid system (22) (pET22b, Novagen). Unlabelled wild-type Ada-C for isothermal titration calorimetry was produced by inoculating 2YT media (supplemented with 100 mg/l ampicillin) with Ada-C expressing cells and allowing growth for 15–17 h prior to harvesting. Induction with isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) was not necessary in this case as leaky expression was observed in 2YT media.

Deuterated and <sup>15</sup>N- or <sup>15</sup>N/<sup>13</sup>C-labelled Ada-C was produced by cell growth in 250 ml of  $2 \times M9$  minimal media made up with 100% <sup>2</sup>H<sub>2</sub>O in place of H<sub>2</sub>O, containing 100 mg/l ampicillin. L-methionine (250 mg/l) was included because B834(DE3) is a methionine auxotroph.

The culture was allowed to grow overnight and the cells were collected by centrifugation. The cell pellet was resuspended into the same growth medium, but containing  $(^{15}NH_4)_2SO_4$  as a sole nitrogen source (for  $^{15}N$ -labelled Ada-C) or both  $(^{15}NH_4)_2SO_4$  and  $^{13}C$ -labelled glucose (for  $^{13}C/^{15}N$ -labelled Ada-C). Expression of Ada-C was induced after 1 h by the addition of IPTG to 1 mM, and allowed to proceed for 4 h before harvesting.

Ada-C selectively labelled with  ${}^{13}C/{}^{15}N$ -cysteine was prepared by inoculating the clone-containing B834(DE3) cells from a plate into 250 ml of 2× M9 minimal media (100% H<sub>2</sub>O) containing defined quantities of each unlabelled amino acid, as outlined by Muchmore *et al.* (23). Following overnight growth the cells were pelleted and resuspended in fresh media identical in all respects to the former media, except for the replacement of unlabelled cysteine with  ${}^{13}C/{}^{15}N$ -labelled cysteine. As indicated above, expression was induced by 1 mM IPTG and allowed to continue for 4 h prior to harvesting and purification.

Following cell disruption by French press, Ada-C was initially purified using an SP Sepharose (Pharmacia) column, eluting with a linear gradient of 0–0.6 M NaCl in 50 mM sodium phosphate buffer pH 7.8, 5 mM DTT and 1 mM

EDTA. This was followed by size-exclusion chromatography (Pharmacia Superdex 75, 50 mM sodium phosphate buffer, pH 7.8, 5 mM DTT, 1 mM EDTA, 400 mM NaCl). Final concentration of the protein to millimolar concentrations for NMR work was achieved through the use of Amicon pressure cells and Centricon concentrators (Amicon Co.). Protein samples were >95% pure as assessed by SDS–PAGE. Electrospray mass spectrometry and N-terminal sequencing analysis were performed to verify the molecular mass, protein sequence and level of deuteration (if applicable). The molecular mass of unlabelled wild-type Ada-C was 19 442 Da. Purified wild-type Ada-C showed 90% specific activity as assayed using established methods (24).

All oligonucleotides for binding studies were 20mer and synthesised on a 1  $\mu$ M scale on an Applied Biosystems 394 automated DNA synthesiser using phosphoramidite chemistry. The methylated oligonucleotide was 5'-AGGTTGC-CACA( $O^6$ meG)GTTGCCAC. The modified phosphoramidite for incorporation of  $O^6$ meG was 5'-dimethyloxytrityl-*N*-isobutyryl- $O^6$ -methyl-2'-deoxyguanosine,3'-[(2-cyanoethyl)-*N*,*N*-diisopropyl]-phosphoramidite (Glen Research).

In the unmethylated oligonucleotide, *O*<sup>6</sup>meG was replaced by guanine. The complementary strand was also synthesised for the creation of duplex DNA. The latter was produced by mixing equimolar amounts of two complementary strands together, heating at 80°C for 5 min and then cooling slowly to room temperature. Annealing was confirmed by PAGE analysis and NMR.

#### Isothermal titration calorimetry

Calorimetric measurements were carried out using an isothermal titration microcalorimeter (MicroCal Inc.). To obtain binding constants, enthalpies and stoichiometries of DNA binding, pure wild-type Ada-C (0.6-0.7 mM) was titrated into unmethylated single- or double-stranded 20mer DNA (0.06-0.08 mM). To gauge protonation effects, titrations were carried out in buffers with different enthalpies of ionisation. Four different buffer systems were used: 50 mM HEPES, MOPS, Tris or phosphate pH 7.8, each containing 5 mM DTT and NaCl at concentrations ranging from 30 to 78 mM to maintain a constant ionic strength (measured as that which produced a conductivity of 9.8 mS at 25°C). Ada-C and DNA solutions were dialysed extensively against each buffer, and the final dialysate was used for concentration adjustments and control titrations into buffer. Typically, Ada-C was injected into the DNA using 20 injections of 12.5 µl each and a delay of 210 s between successive injections. The cell (1.34 ml working volume) was stirred at a rate of 400 r.p.m. All experiments were conducted at 25°C. Just prior to a titration experiment, all solutions were degassed using the Thermolyne degasser supplied with the ITC unit. Data analysis to determine the thermodynamic parameters of Ada-C-DNA binding was carried out using the Origin software package (MicroCal Inc.) A non-linear iterative Marquardt method was used for data fitting and a single site model was assumed. The heat of dilution of Ada-C into buffer was negligible.

## NMR assignment of the Ada-C backbone

All NMR experiments were carried out on a Bruker DRX 600 spectrometer at 25°C (Table 1). Backbone assignment of wild-type Ada-C was performed using a 1.3 mM sample of 70% deuterated,  $^{15}N/^{13}C$ -labelled Ada-C in 50 mM sodium

phosphate buffer pH 6.7 (90%  $H_2O$ , 10%  $^2H_2O$ ), 10 mM  $^2H$ -DTT, 1 mM  $^2H$ -EDTA, 100 mM NaCl and 100mM NaBr. Each triple resonance experiment was carried out in the gradient-enhanced, sensitivity-enhanced mode. In each case, 16 scans were performed and the data were processed using Felix 97 (Biosym Inc.) with a 65° phase-shifted sine-bell window function prior to Fourier transformation with linear prediction in the nitrogen frequency dimension.

 Table 1. The triple resonance experiments performed

 on 70% deuterated, <sup>13</sup>C/<sup>15</sup>N-labelled Ada-C, 1.3 mM at

 298 K and 600 MHz field strength

Experiment	Time domain sizes (points)			Spectral widths (Hz)		
	t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	ω1	ω2	ω3
HNCO	442	50	54	77	2272	2212
	(H)	(C)	(N)	88		
HN(CA)CO	442	50	54	77	2272	2212
	(H)	(C)	(N)	88		
HN(CO)CA	442	48	54	83	4200	2212
	(H)	(C)	(N)	88		
HNCA	442	50	54	83	4200	2212
	(H)	(C)	(N)	88		
HN(COCA)CB	442	64	54	77	9600	2212
	(H)	(C)	(N)	88		
HN(CA)CB	442	64	54	77	9600	2212
IIII(CA)CD	(H)	(C)	(N)	88		2212

All points shown are complex points. The number of complex points in all processed spectra were  $t_1 = 512$ ,  $t_2 = 256$  and  $t_3 = 256$ .

A semi-automated assignment strategy was used with an inhouse programme based on simulated annealing. <sup>1</sup>H chemical shifts are expressed relative to the methyl resonance of sodium 2,2-dimethyl-2-silapentane-5-sulphonate at 0 p.p.m.

Following final backbone assignment, chemical shift indexing of the carbon frequencies was performed using the method and software developed and described by Wishart and Sykes (25).

## **DNA-binding analysis by NMR**

DNA binding was monitored by recording <sup>1</sup>H-<sup>15</sup>N HSQC spectra of Ada-C in the presence of ss- or dsDNA. A known amount of solid DNA was dissolved in a small volume of water and the pH adjusted to 6.7 using 0.1 M NaOH. The DNA was then lyophilised and finally dissolved in Ada-C solution to obtain the required ratio of protein:DNA. The pH was verified by direct meter reading.

A high salt NMR buffer (50 mM sodium phosphate, pH 6.7, 10 mM <sup>2</sup>H-DTT, 1 mM <sup>2</sup>H-EDTA, 100 mM NaCl, 100 mM NaBr, 90%  $H_2O/10\%$  <sup>2</sup> $H_2O$ ) was used to obtain weak binding. The protein–DNA complex was in fast exchange under the NMR timescale, hence allowing the determination of an approximate binding constant between Ada-C and single-stranded 20mer DNA by NMR titration.

A lower salt NMR buffer (50 mM sodium phosphate, pH 6.7, 10 mM  $^{2}$ H-DTT, 1 mM  $^{2}$ H-EDTA, 10 mM NaCl, 90% H<sub>2</sub>O/10%  $^{2}$ H<sub>2</sub>O) was used to enhance the binding affinity and increase the degree of complex formation.

<sup>15</sup>N-labelled, 70% deuterated inactive C146S Ada-C was also complexed with both double- and single-stranded methylated and unmethylated DNA for a comparison of binding modes.

Demethylation reactions were also carried out *in situ* in the NMR tube by the addition of ssDNA containing an *O*<sup>6</sup>meG to <sup>15</sup>N-labelled, 70% deuterated wild-type Ada-C in various ratios. The effect on the Ada-C structure of methyl group transfer to the active site cysteine was then examined by NMR. Methyl group transfer from DNA to Ada-C was confirmed by electrospray mass spectrometry on a Micromass Platform electrospray mass spectrometer.

The tautomeric state of the imidazole sidechain of His147 was determined using <sup>1</sup>H-<sup>15</sup>N HMQC experiments in the manner of Damblon *et al.* (26) for both free and DNA-bound wild-type Ada-C.

A pH titration to determine the  $pK_a$  of Cys146 was carried out on wild-type Ada-C labelled with  ${}^{13}C/{}^{15}N$ -cysteine in 100%  ${}^{2}H_2O$  high salt NMR buffer.  ${}^{1}H{}^{-13}C$  HSQC spectra were recorded at 600 MHz and 298 K for Ada-C samples at pH values uncorrected for deuterium (pH 6.3–11).

# Modelling DNA binding

The Dali program (27,28) was used to identify structural homology between Ada-C and other DNA-binding proteins.

By combining structural alignment data with the mapped DNA-binding site, a model of how Ada-C binds and repairs O<sup>6</sup>meG-containing DNA was devised. The DNA coordinates used for the model were those from the structure of the HhaI methyltransferase-DNA complex (29). The 'flipped out' (extrahelical) cytosine was mutated to an O<sup>6</sup>meG using the Builder module in Insight II (MSI Inc.). Following docking of Ada-C onto the DNA, energy minimisation was performed on the region within 8 Å of the C5 atom of the extrahelical guanine using the Macromodel/Batchmin package (v5.5) (30). Within this region, different restraining potentials were applied to parts of protein and DNA. Unless otherwise stated, a force restraint of 100 kJ/Å was applied to the atoms to constrain them to their crystallographic coordinates. The sidechains of residues Y115, V128 and L140-V150 were left without force restraint. So too was the extrahelical guanine base, and a segment of the DNA backbone, which comprised the 3 nt 5' of this guanine and nucleotides A405, G406 and C407 (numbering from PDB file 1MHT) on the complementary DNA strand. The unpaired cytosine base C408, opposite the extrahelical guanine, was constrained by 50 kJ/Å. The protein backbone of residues Y115 and I142-V149 were constrained by 25 kJ/Å. Energy minimisation was performed using the Steepest Descent Gradient Procedure, and was terminated after 1000 iterations. Minimisation was then continued using a PRCG procedure which was terminated after the energy gradient r.m.s. fell below 0.01 kJ/mol Å.

# RESULTS

## The DNA-binding site on Ada-C

A disadvantage of the triple-labelling procedure used was that methionine residues were not <sup>15</sup>N- or <sup>13</sup>C-labelled, and thus could not be assigned. This was the sacrifice made for Ada-C yield gains by using B834(DE3).



**Figure 1.** Sequence and secondary structure of Ada-C. α-helices are depicted as rectangles, β-strands as arrows, helix–turn–helix motif helices 4 and 5 are shown in grey. In the presence of a 2.5-fold excess of single-stranded unmethylated 20mer DNA certain chemical shift changes were observed for Ada-C resonances. These were grouped into two categories. (i) Chemical shifts which changed, and did so by <0.15 p.p.m. in <sup>1</sup>H-dimension, or <0.5 p.p.m. in <sup>15</sup>N-dimension in high salt buffer and shifted but did not broaden significantly in low salt buffer (coloured blue). (ii) Chemical shifts which changed by >0.15 p.p.m. in <sup>1</sup>H-dimension in high salt buffer, or whose signals were extensively line-broadened, such that they disappeared upon binding in low salt buffer (coloured red). This figure represents the chemical shift map of the DNA-binding face of Ada-C. This chemical shift mapping colour scheme is also depicted on the protein itself in Figure 5.

Despite this, triple resonance experiments on 70% deuterated <sup>13</sup>C/<sup>15</sup>N-labelled Ada-C resulted in the backbone assignment of 96% of amino acid residues. These assignments allowed chemical shift changes upon DNA binding to be mapped to specific regions of the protein backbone. Similar chemical shift mapping strategies have been widely used to localise intermolecular binding surfaces in other DNA/protein systems (31-33). In the present case, binding of Ada-C to either ss- or dsDNA involved a similar pattern of backbone amide chemical shift changes. This implies that Ada-C uses the same binding surface in each case, and that binding to duplex DNA occurs predominantly to a single DNA strand. Chemical shift movements were placed into two categories: those involving less than, or greater than, 0.15 p.p.m. changes in proton and 0.5 p.p.m. change in <sup>15</sup>N dimensions, upon complex formation of Ada-C with a 2.5-fold excess of DNA in high ionic strength buffer (Fig. 1). The latter category of chemical shifts were subject to line broadening and signal loss, upon complex formation in low ionic strength buffer. This selective linebroadening occurred for resonances from the recognition helix of the helix-turn-helix (helix 5), and the loop region made up of residues 151-160. Together, these structures form a helixturn-wing motif (20), which forms the DNA-binding interface of Ada-C. The switch from high to low salt buffer resulted in substitution of a fast chemical exchange between free and complexed protein, for one of intermediate exchange, typically occurring on the millisecond timescale (34,35). Intermediate exchange line-broadening for resonance signals emanating from the recognition helices of helix-turn-helix motifs is a recurrent theme in protein-DNA complexes (36-38), and has also been noted to occur with Ada-N (39). Such localised linebroadening has been attributed to flexibility within the protein-DNA interface (40). These observations enforce the notion that DNA recognition is a highly dynamic process.

Protein flexibility at the DNA-binding site would render the protein more amenable to enveloping an extrahelical base than rigid docking during base flipping. Conformational fluctuations in the binding site may also confer an entropic advantage to the binding interaction (35).

Upon binding single- or double-stranded unmethylated DNA, HMQC experiments revealed an unaltered peak pattern, revealing that the Nɛ2H tautomeric state of His147 was maintained upon binding (spectra not shown). This ensures that the imidazole proton is shared through a hydrogen bond with Glu173, as previously postulated (41).

The exact  $pK_a$  of active site residue Cys146 could not be determined as the protein denatured at pD values >10. No titration of the C $\alpha$  or C $\beta$  chemical shift was observed up to this pD value (spectra not shown), and the conclusion drawn is that Cys146 has a  $pK_a$ >9.

#### Reaction of Ada-C with methylated DNA

In vitro methylation reactions were monitored by NMR. Titration of 20mer oligonucleotide containing an O<sup>6</sup>meG into wild-type Ada-C caused subsequent precipitation of the protein in low ionic strength buffer. Electrospray mass spectrometry on both DNA and Ada-C revealed that methyl transfer had occurred and confirmed the methylated state of the unfolded protein. Precipitation of methylated Ada-C was avoided by first employing high ionic strength buffer, and secondly making the addition of excess O<sup>6</sup>meG-containing 20mer DNA to Ada-C in the presence of equimolar unmethylated 20mer DNA. A comparison between <sup>1</sup>H-<sup>15</sup>N HSQC spectra of methylated and unmethylated Ada-C in complex with ssDNA revealed that resonances from the DNA-binding subdomain (residues 93-177) were extensively broadened upon methylation, whilst those from the N-terminal subdomain remained unaffected. This suggested that a localised unstructuring occurred in the DNAbinding subdomain as a result of self-methylation. Mutation of the active site cysteine to an alanine resulted in a protein which precipitated when concentrated to 0.1 mM. Alteration of the same cysteine to a serine, however, resulted in a protein which was stable at millimolar concentrations. The <sup>1</sup>H-<sup>15</sup>N HSOC spectra of C146S mutant Ada-C contained resonances with broad linewidths, suggesting increased conformational flexibility relative to wild-type Ada-C.

Taken together, these results reinforce previous findings that disruption of the active site hydrogen bonding network (of which the Cys146 sulphydryl forms a part) leads to structural instability (42). Methylation of Cys146 resulted in a reduction in structural stability which was localised to the DNA-binding subdomain of Ada-C (residues 93–177, containing the mapped DNA-binding site). The structural integrity of the N-terminal subdomain (residues 1–92, consisting of the  $\beta$ -sheet and two  $\alpha$ -helices one and two) was unaffected by methylation of Cys146.

Using a 70% deuterated, <sup>15</sup>N-labelled sample of mutant C146S Ada-C, the DNA-binding site for methylated and unmethylated duplex DNA was mapped to the helix–turn–wing motif, as in the wild-type protein. This implies that this motif is used when binding both methylated and unmethylated duplex DNA. Qualitatively, based on <sup>1</sup>H and <sup>15</sup>N chemical shift changes, Ada-C C146S appears to bind with greater affinity to DNA containing *O*<sup>6</sup>meG than to unmethylated DNA.

**Figure 2.** <sup>1</sup>H-<sup>15</sup>N HSQC spectra of 70% deuterated, <sup>15</sup>N-labelled Ada-C C146S mutant in complex with equimolar methylated duplex DNA (red) and unmethylated duplex DNA (blue). Both spectra were recorded at 600 MHz and 298 K, pH 6.7 in low salt buffer. The complex of Ada-C with methylated DNA appeared to be in a slower exchange regime than the complex with unmethylated DNA.

The differences observed in the <sup>1</sup>H-<sup>15</sup>N HSQC spectra of this protein bound to methylated and unmethylated DNA may result from Ada-C flipping *O*<sup>6</sup>meG but not unmodified bases (Fig. 2).

#### **Energetics of DNA binding and demethylation**

On the basis of isothermal titration calorimetry, the affinity of wild-type Ada-C for unmethylated 20mer single-stranded and duplex DNA was similar ( $K_d = 2 \mu M$ ). This dissociation constant corresponds well with the binding constant determined for whole Ada by anisotropic fluorescence studies (43). Two molecules of Ada-C bound to 20mer dsDNA whilst only one Ada-C molecule bound single-stranded 20mer DNA. The energetics profile of binding was identical for single-stranded and duplex DNA, consistent with the suggestion of similar binding modes. An endothermic  $\Delta H$  ( $\Delta H = +0.53$  kcal/mol) but favourable  $\Delta S$  ( $T\Delta S = +8.3$  kcal/mol) indicates that DNA binding is entropically driven (Figs 3 and 4). We suggest that upon binding, ordered ions and water molecules from the protein and DNA surfaces are lost to bulk solution, providing the increase in entropy.

To assess the energetics of demethylation, Ada-C was titrated into a large excess of ssDNA containing  $O^6$ meG (a 300-fold excess of DNA was present after the first injection of Ada-C). Using this vast excess of DNA ensured an enthalpy change of approximately equal magnitude with each injection of Ada-C. This measured enthalpy change was strongly exothermic and highly buffer-dependent (Table 2). The magnitude of this

**Figure 3.** The raw and processed (integrated) data for the titration of 0.72 mM Ada-C into unmethylated, single-stranded 0.06 mM DNA in phosphate buffer. The fit to a single-site binding model is given as the sigmoidal curve.

3.5

3

2

1.5

0.5

+ 0 0

AHmeas (kcal/mol)

y = 0.2266x + 0.5311



6

8

 $\Delta H_{ion}$  (kcal/mol)

10

12

enthalpy change is a combination of the energy changes involved in binding Ada-C to methylated DNA, the demethylation reaction, subsequent structural alterations of Ada-C, as well as any buffer ionisation effects. One or more of these enthalpies may be buffer dependent, leading to the different enthalpy magnitudes observed. From these results it appears



Time (min) 30 50 60 40 70 80 0.2 ucal/sec 0.1 0.0 -0.1 0.8 injectan **6** 0.4 kcal/mole 0.2 0.0 0.0 0.5 1.0 1.5 2.0 2.5 Molar Ratio

ppn

106

110

likely that the previously proposed  $S_N^2$  demethylation reaction (6,44,45) is highly exothermic.

**Table 2.** The approximate heats of reaction for demethylation of methylated ssDNA by Ada-C in four buffers

Buffer	Heat of reaction for				
	demethylation of ssDNA (kcal/mol)				
Phosphate	-41				
Tris	-33				
MOPS	-25				
HEPES	-29				

These  $\Delta H$  values were obtained by integrating the area defined by the exothermic peaks resulting from additions of small amounts of Ada-C into 300-fold excess of methylated ssDNA. The values so obtained were then averaged to arrive at the figures shown (since the figures are only approximate the standard errors were not included). Contributions to the observed enthalpy change are made not only from the demethylation reaction, but also by partial unfolding of Ada-C following methylation, the initial binding reaction and any buffer ionisation effects. Each of these contributions may be buffer dependent, leading to the observed overall heat of reaction shown in the table also being highly buffer dependent.

## DNA-binding model of Ada-C

Localisation of the DNA-binding interface to the helix-turn-wing motif enabled a more accurate DNA-binding model to be constructed. Using the Dali structural alignment algorithm (46) we identified other proteins containing structurally homologous DNA-binding motifs. Of those identified by the search, the protein with the best characterised DNA-binding mode was the E.coli catabolite activator protein (CAP) as previously identified by Daniels et al. (19) and Wibley et al. (47), using similar methods. The helix-turn-helix and wing region of Ada-C (residues Y115 to R160) showed a 3.3 Å r.m.s.d. compared to the  $\alpha$ -carbons from the helix-turn-helix region in CAP. The structure of the CAP/DNA complex (48) (PDB ID: 2CGP) was used as a template on which to base a model of the Ada-C-DNA complex. The helix-turn-helix regions from Ada-C (helices 3, 4 and 5) and CAP (helices 5, 6 and 7) were superimposed manually. The CAP was then removed, leaving Ada-C bound to bent CAP operator DNA, with the recognition helix lying in the major groove. In this DNA conformation any base lesions stacked within the duplex would be inaccessible to the buried active site Cys146, the methyl acceptor of Ada-C. The current work revealed that the postulated gross conformational changes in the structure of Ada-C (41) do not occur upon repair of methylated DNA. In order to gain accessibility to the active site, the DNA must instead undergo a conformational change. It is now widely believed that Ada-C and homologous alkyltransferases act on extrahelical base lesions (19). Such a lesion, 'flipped out' of duplex DNA (rendered extrahelical), could fit through an opening in the surface of Ada-C between the 'wing' and the recognition helix to lie in the buried active site.

To obtain such a productive binding model, the CAP operator DNA was substituted for the structure of DNA from the *HhaI* methyltransferase–DNA complex structure (29) (PDB ID: 1MHT). The extrahelical cytosine in this DNA was mutated to  $O^6$ meG using the Builder module in Insight II (MSI Inc.). This flipped base lesion was then inserted into the active site of Ada-C

through the opening between the wing and the recognition helix (Fig. 5) and the model energy minimised. This model orientated Ada-C in a manner consistent with the DNA-binding face only making contacts to a single DNA strand, in agreement with both the NMR and ITC data.

# DISCUSSION

The NMR spectra of Ada-C in the presence and absence of DNA were overall the same, with differences limited to localised chemical shift and linewidth alterations, corresponding to residues involved in the DNA-binding site. This suggests that any conformational changes in Ada-C upon binding methylated or unmethylated DNA were minimal. There does not appear to be any movement of the C-terminal helix into the major groove of DNA, contrary to a previous proposal (41). Chemical shift mapping revealed that the DNA-binding interface of Ada-C consisted of a helix–turn–wing motif.

The enthalpy change and dissociation constant were very similar when Ada-C bound duplex or single-stranded unmethylated DNA. A doubling in the stoichiometry of Ada-C was, however, observed upon binding to duplex relative to ssDNA. NMR data demonstrated conclusively that there was no difference in the manner by which Ada-C bound either ssor dsDNA. Taken together, these structural and thermodynamic data suggest that when binding duplex DNA, Ada-C binds predominantly to a single DNA strand.

This appears to be a common theme amongst single-base flipping DNA repair proteins (49). Various DNA glycosylases also only contact the phosphate backbone of a single DNA strand when bound to duplex DNA (49–51). This contrasts to the majority of DNA-binding proteins, which form contacts to both DNA strands.

In our model of the complex, Ada-C contacts a single-strand of duplex DNA. Helix 5 is employed to orientate the protein in the major groove. The conserved Arg129 in this helix is aptly poised to either form a potential salt link to the sugar-phosphate backbone of DNA, or to act as a probe to initiate DNA base flipping, as has been recently suggested (19). The 'wing' (residues 151-160) also forms contacts to the DNA backbone, notably from Arg151. Although this residue is not strictly conserved amongst the family of alkytransferases, there are often other residues nearby which could form a hydrogen bond to the DNA backbone (e.g. Ser151 in human alkyltransferase). Arg151 in Ada-C therefore represents a potential site of interaction with the DNA. In comparison with other DNA-binding proteins the number of salt links and hydrogen bonds formed between protein and DNA is low. This may account for the small endothermic enthalpy change ( $\Delta H$ = +0.53 kcal/mol) observed on DNA binding by Ada-C. Rather than defined enthalpy-based interactions, Ada-C relies on entropic factors to drive the binding to DNA. The most predominant driving force is likely to be the release of ordered water molecules and ions from the surfaces of protein and DNA upon docking. Burial of the base lesion in the hydrophobic active site of Ada-C may also confer an entropic advantage, and perhaps some of the driving force to the base-flipping process.

Once the base is flipped, only a minimal amount of deformation of the 'wing' of Ada-C was required to accommodate the  $O^6$ meG lesion within the active site. We orientated the flipped  $O^6$ meG lesion in the active site in a manner consistent with the



**Figure 5.** DNA-binding model of Ada-C with  $O^6$ meG flipped into the active site through the opening between the recognition helix and wing. The orientation of the recognition helix of Ada-C with respect to the DNA is based on the CAP–DNA complex (47). The colour scheme is identical to that used in Figure 1, and represents the two categories of chemical shift movement observed for Ada-C backbone resonances upon DNA binding. Backbone regions of Ada-C most affected by the DNA binding are coloured red, those less affected are coloured blue. Unaffected regions are coloured grey or white. Ada-C predominantly contacts a single strand of the duplex. Sidechains from residues showing potential interactions to the DNA backbone of the flipped strand are depicted. The exocyclic methyl group from  $O^6$ meG is shown as a blue ball in the active site, juxtaposed with the sidechain of Cys146 (blue).



**Figure 6.** The active site of Ada-C containing extrahelical  $O^6$ meG. This model was generated by orientating the Ada-C sidechains depicted, along with the position of  $O^6$ meG, to demonstrate potential interactions. Potential hydrogen bonds are represented by dotted lines. Note that the exocyclic oxygen of  $O^6$ meG occupies the same space as an ordered water molecule (shown as blue spheres). The sulphur of Cys146 is juxtaposed with the methyl group of  $O^6$ meG. This figure was generated using Raster3D (60).

studies performed on the repair of analogues (5–7) (Fig. 6). Potential hydrogen bonds from the sidechains of conserved K166 to either  $O^6$  or  $N^1$  of the base, and conserved Y115 to the  $N^3$  or  $N^2$  amino group of the base may function to orientate

 $O^6$ meG productively in the active site. The Nɛ2H tautomeric state of His147, and consequently the hydrogen-bonding network of the catalytic tetrad, Glu173–His147–water–Cys146 is undisrupted by entry of the base into the active site.

NMR pH titrations revealed that in the free form of the protein, the thiol of Cys146 did not have an abnormally low  $pK_{a}$  (i.e. <9). Based on these observations we speculate that deprotonation of the thiol occurs concurrently with nucleophilic attack by the sulphur atom on the methyl group of the base. Following demethylation of the  $O^6$ meG, a build up of negative charge on the base would have to be stabilised. This could be partially achieved by resonance delocalisation around the aromatic ring of the base. The positively charged sidechain of the conserved Lys166 is also perfectly located to aid in stabilising this negative charge until protonation of the O<sup>6</sup> is achieved. Both Lys166 and Tyr115 are suitably located to act as proton donors. However, their sidechain  $pK_a$  values would have to be significantly lower than the  $pK_a$  of free residues  $(pK_a \approx 10)$ . Alternatively, water might act as a general acid and reprotonate the base.

Upon methylation of Cys146 (shown above to be a strongly exothermic reaction), the DNA-binding subdomain of Ada-C loses structural integrity. An identical phenomenon was recently observed upon the alkylation of the analogous cysteine in the human homologue (19). This instability is thought to assist in dissociation from the DNA, and proteolytic degradation. As a 'suicide' enzyme which acts only once, Ada-C has not evolved for persistence following DNA repair.

The model presented here is very similar to one previously postulated by Vora *et al.* (18). The only difference is the angle at which helix 5 lies across the major groove of DNA; a detail which will have to await the elucidation of a high resolution structure of the complex. The Vora model also conforms well to the experimentally mapped DNA-binding surface of Ada-C.

Although flipping of the O<sup>6</sup>meG lesion has not been experimentally proven, it seems to be a functional necessity in order to explain the very minor changes of the Ada-C conformation upon DNA binding. Arg129 seems a likely candidate for filling the hole in the base stack left by rendering a lesion extrahelical, and perhaps even forming hydrogen bonds with the orphan cytosine. It has been suggested that the analogous R128 in human alkyltransferase acts to actively extrude the lesion in an analogous manner to the leucine on the minor-groove binding loop of uracil DNA-glycosylase (UDG) (19). In UDG, the minor groove binding loop pushes the uracil base out of the major groove (52). It is hard to see how R129 in Ada-C could act in an analogous manner since it lies in the major groove of DNA. Perhaps it acts by 'pulling' out the lesion from the DNA base stack. Alternatively, R129 might simply block the lesion's return to the DNA base stack through filling the hole left behind by the flipped base. Initially some strain might be placed on the DNA backbone by interactions from the 'wing' and the recognition helix of Ada-C. Perhaps this action, analogous to the 'pinch' of the UDG on its target DNA backbone, and the 'pull' of the lesion by forming complementary interactions within the active site, is sufficient to dislocate the lesion from the base stack. It seems unlikely that the highly efficient repair capabilities of the alkyltransferases rely on serendipidous capture of spontaneously flipped out O<sup>6</sup>meG lesions.

Recognition of the  $O^6$ meG lesions amidst the mass of competing genomic DNA is a mammoth task. A DNA-repair protein must either localise the lesion by direct contact, indirectly via some perturbation to the DNA structure, or must individually sample all bases (by extrahelical extrusion) until a lesion is located. The one-dimensional sliding along the DNA, as thought to occur for the *Lac* repressor (53), is believed not to occur with human alkyltransferase (54). Rather, Ada-C moves along the DNA using a series of minicollisions, between which it can translocate along the DNA backbone. This should be facilitated by the protein's charge dipole which would enable electrostatic orientation of the DNA-binding face towards the phosphate backbone of DNA. Ada-C would then be aptly suited to scanning for lesions whilst hopping on and off the DNA.

Lesion localisation would potentially be more rapid if Ada-C did not flip every base to sample it, as has been suggested for glycosylases (55). Observed differences in the protein spectra of inactive C146S Ada-C bound to methylated and unmethylated DNA might be indicative of base flipping in the case of methylated but not unmethylated DNA. Perhaps Ada-C only flips the alkylated base, an action which may be facilitated by the localised lesion-induced weakening of the stacked structure.

 $O^6$ meG is thought to promote localised unstacking in the DNA duplex, which may in turn lead to slight kinking of the structure (56,57). In this manner a lesion could advertise its

presence more dramatically to DNA repair proteins. A mechanism whereby Ada-C locates the lesion via the slight deformation caused to the DNA structure would explain why ssDNA is repaired more slowly than duplex DNA, even though the two types are bound with similar affinity (58,59). A lesion in the less structured ssDNA might be less readily located than in B-form duplex DNA, because there is less induced structural perturbation in the former.

In summary, we have obtained direct experimental evidence to show that Ada-C binds to a single strand of the dsDNA, and that the protein undergoes minimal conformational change upon binding methylated DNA. The latter observation supports the hypothesis that base flipping is required for Ada-C to perform its demethylation function.

## ACKNOWLEDGEMENTS

We would like to thank the University of Leicester for the research scholarship granted to P.E.V., Drs J. Eads and H. Trayer (University of Birmingham, UK) for the use of their ITC apparatus, Dr G. Major and Ms S. Charlton (University of Newcastle, UK) for assaying Ada-C, Professor S. Ealick (Cornell, USA) for providing the coordinates of the model reported in (18), Drs K. Lilley and S. Mistry (PNACL, Leicester, UK) for synthesising all oligonucleotides and Dr I. Barsukov (Leicester, UK) for help with the NMR. Figure 6 was made using the program Prepi, kindly made available by Dr S. Islam at the Imperial Cancer Research Fund. PDB coordinates of the model discussed here are freely available.

### REFERENCES

- 1. Olsson, M. and Lindahl, T. (1980) J. Biol. Chem., 255, 10569-10571.
- 2. Lindahl, T., Demple, B. and Robins, P. (1982) EMBO J., 1, 1359-1363.
- 3. Teo,I., Sedgwick,B., Demple,B., Li,B. and Lindahl,T. (1984) *EMBO J.*, **3**, 2151–2157.
- 4. Demple, B., Sedgwick, B., Robins, P., Totty, N., Waterfield, M.D. and Lindahl, T. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 2688–2692.
- 5. Spratt, T.E. and DeLosSantos, H. (1992) Biochemistry, 31, 3688-3694.
- 6. Spratt, T.E. and Cambell, C.R. (1994) Biochemistry, 33, 11364-11871.
- 7. Spratt, T.E., Wu, J.D., Levy, D.E., Kanugula, S. and Pegg, A.E. (1999) Biochemistry, **38**, 6801–6806.
- Teo, I., Sedgwick, B., Kilpatrick, M.W., McCarthy, T.V. and Lindahl, T. (1986) Cell, 45, 315–324.
- Nakamura, T., Tokumoto, Y., Sakumi, K., Koike, G., Nakabeppu, Y. and Sekiguchi, M. (1988) J. Mol. Biol., 202, 483–494.
- Akimaru,H., Sakumi,K., Yoshikai,T., Anai,M. and Sekiguchi,M. (1990) J. Mol. Biol., 216, 261–273.
- 11. Saget, B.M. and Walker, G.C. (1994) Proc. Natl Acad. Sci. USA, 91, 9730–9734.
- 12. Landini, P., Bown, J.A., Volkert, M.R. and Busby, S.J.W. (1998) *J. Biol. Chem.*, **273**, 13307–13312.
- 13. Sakumi, K. and Sekiguchi, M. (1989) J. Mol. Biol., 205, 373–385.
- Lindahl, T., Sedgwick, B., Sekiguchi, M. and Nakabeppu, Y. (1988) Annu. Rev. Biochem., 57, 133–157.
- Verdemato, P.E. and Moody, P.C.E. (1998) Nucleic Acids Mol. Biol., 12, 1–27.
- 16. Moody,P.C.E. and Moore,M.H. (1995) Contrib. Oncol. Basel Karger, 49, 16–24.
- Kanugula,S., Goodtzova,K., Edara,S. and Pegg,A.E. (1995) *Biochemistry*, 34, 7113–7119.
- Vora, R.A., Pegg, A.E. and Ealick, S.E. (1998) *Proteins Struct. Funct. Genet.*, **32**, 3–6.
- Daniels,D.S., Mol,C.D., Arvai,A.S., Kanugula,S., Pegg,A.E. and Tainer,J.A. (2000) *EMBO J.*, **19**, 1719–1730.
- 20. Brennan, R.G. (1993) Cell, 74, 773-776.
- 21. Sayers, J.R., Krekel, C. and Eckstein, F. (1992) Biotechniques, 13, 592-596.

- Doherty, A.J., Ashford, S.R., Brannigan, J.A. and Wigley, D.B. (1995) Nucleic Acids Res., 23, 2074–2075.
- 23. Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, E. and
- Dahlquist, F.W. (1989) Methods Enzymol., 177, 44-73.
- 24. Major,G.N., Gardner,E.J. and Lawley,P.D. (1991) *Biochem. J.*, **277**, 89–96.
- 25. Wishart,D.S. and Sykes,B.D. (1994) *J. Biol. NMR*, **4**, 171–180.
- Damblon, C., Prosperi, C., Lian, L.-Y., Barsukov, I., Soto, R.P., Galleni, M., Frere, J.-M. and Roberts, G.C.K. (1999) J. Am. Chem. Soc., 121, 11575–11576.
- 27. Holm,L. and Sander,C. (1994) *Proteins*, **19**, 165–173.
- Holm,S. and Sander,C. (1996) *Science*, 273, 595–602.
   Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) *Cell*, 76,
- Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R.M., Lipton, M.,
- Caufield, C., Chang, G., Hendrickson, T. and Still, W.C. (1990) *J. Comput. Chem.*, **11**, 440–467.
- Artz,P.G., Valentine,K.G., Opella,S.J. and Lu,P. (1996) J. Mol. Recogn., 9, 13–22.
- Foster, M.P., Wuttke, D.S., Clemens, K.R., Jahnke, W., Radhakrishnan, I., Tennant, L., Reymond, M., Chung, J. and Wright, P.E. (1998) *J. Biol. NMR*, 12, 51–71.
- Holmbeck,S.M.A., Dyson,H.J. and Wright,P.E. (1998) J. Mol. Biol., 284, 533–539.
- Kay,L.E., Muhandiram,D.R., Farrow,N.A., Aubin,Y. and Forman-Kay,J.D. (1996) *Biochemistry*, 35, 361–368.
- Foster, M.P., Wuttke, D.S., Radhakrishnan, I., Case, D.A., Gottesfeld, J.M. and Wright, P.E. (1997) *Nature Struct. Biol.*, 4, 605–608.
- Dekker, N., Cox, M., Boelens, R., Verrijzer, C.P., Vandervliet, P. and Kaptein, R. (1993) *Nature*, 362, 852–855.
- Ramesh, V., Frederick, R.O., Syed, S.E.H., Gibson, C.F., Yang, J.C. and Roberts, G.C.K. (1994) *Eur. J. Biochem.*, 225, 601–608.
- 38. Vuister, G.W., Kim, S.J., Wu, C. and Bax, A. (1994) Biochemistry, 33, 10-16.
- Sakashita,H., Sakuma,T., Ohkubo,T., Akitomo,Y., Kainosho,M., Sekiguchi,M. and Morikawa,K. (1995) J. Biochem., 118, 1184–1191.

- Slijper, M., Boelens, R., Davis, A.L., Konings, R.N.H., van der Marel, G.A., van Boom, J.H. and Kaptein, R. (1997) *Biochemistry*, 36, 249–254.
- Moore, M.H., Gulbis, J.M., Dodson, E.J., Demple, B. and Moody, P.C.E. (1994) *EMBO J.*, **13**, 1495–1501.
- 42. Crone, T.M., Goodtzova, K. and Pegg, A.E. (1996) Mutat. Res., 363, 15-25.
- Takahashi, M., Sakumi, K. and Sekiguchi, M. (1990) *Biochemistry*, 29, 3431–3436.
- Morimoto, K., Dolan, M.E., Scicchitano, D. and Pegg, A.E. (1985) Carcinogenesis, 6, 1027–1031.
- 45. Graves, R.J., Li, B.F.L. and Swann, P.F. (1989) Carcinogenesis, 10, 661-666.
- 46. Holm,L. and Sander,C. (1993) J. Mol. Biol., 233, 123-138.
- Wibley, J.E.A., Pegg, A.E. and Moody, P.C.E. (2000) Nucleic Acids Res., 28, 393–401.
- 48. Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) Science, 253, 1001–1007.
- 49. Bruner, S.D., Norman, D.P.G. and Verdine, G.L. (2000) Nature, 403, 859-866.
- 50. Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E. and Tainer, J.A. (1998) *EMBO J.*, **17**, 5214–5226.
- Barrett, T.E., Scharer, O.D., Saava, R., Brown, T., Jiricny, J., Verdine, G.L. and Pearl, L.H. (1999) *EMBO J.*, 18, 6299–6609.
- 52. Slupphaug,G., Mol,C.G., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) *Nature*, **384**, 87–92.
- 53. von Hippel, P.H. and Berg, O.G. (1989) *J. Biol. Chem.*, **264**, 675–678. 54. Bender, K., Federwisch, M., Loggen, U., Nehls, P. and Rajewsky, M.F.
- (1996) Nucleic Acids Res., **24**, 2087–2094.
- 55. Verdine, G.L. and Bruner, S.D. (1997) Chem. Biol., 4, 329–334.
- 56. Voigt, J.M. and Topal, M.D. (1990) Biochemistry, 29, 5012–5018.
- 57. Wong, C.-W., Tan, N.-W. and Li, B.F.L. (1992) J. Mol. Biol., 228, 1137–1146.
- Bhattacharyya, D., Foote, R.S., Boulden, A.M. and Mitra, S. (1990) Eur. J. Biochem., 193, 337–343.
- Fried,M.G., Kanugula,S., Bromberg,J.L. and Pegg,A.E. (1996) Biochemistry, 35, 15295–15301.
- 60. Merrit, E.A. and Murphy, M.E. (1994) Acta Crystallogr., D50, 869-873.